# Risk and Failure Probability, What is it?

## Terminology

1. **Risk** is the:
   a. Potential of losing something of value (e.g., life, property, performance, schedule, or cost).
   b. Effect of uncertainty on objectives. (Ref. ISO 31000, 2009)
2. A **risk statement** contains three elements (e.g., as in three columns in a table), namely:
   a. Scenario, what can go wrong?
   b. Likelihood, what is the probability it will happen?
   c. Consequence, what is the impact if it did happen?
3. **Reliability** is the:
   a. Likelihood
   b. An item will perform its intended function
   c. For a stated mission time
   d. Under stated conditions.

## Risk vs. Reliability

1. The probability used in:
   a. Risk is the **probability of failure**, denoted $\boldsymbol{Pf}$, for the item of interest.
   b. Reliability is the **probability of success**, denoted $\boldsymbol{Ps}$, for the item of interest.
2. Fundamental math rule: $\boldsymbol{Pf + Ps = 1}$.
3. When one type of probability is known, the other type can be easily determined by its complement.
4. Furthermore, the complement of the reliability measure makes the **likelihood axis of the risk matrix**, and the complement of safety makes the **consequence axis of the risk matrix**.

## Types of Data and Methods Commonly Used to Make a Probability of Failure

1. **Demand-based (pass-fail events)**: For example, item x (e.g., starter solenoid) successfully completed its mission upon demand. A data set consisting of items with this type of **discrete** life data with independent trials is often modeled with the *binomial* distribution using the probability of failure ($p$) where:
   a. $p = \frac{failure\ count}{total\ number\ of\ attempts}$ based on classical statistics.
   b. $p = \frac{failure\ count+0.5}{total\ number\ of\ attempts+1}$ based on one version of Bayesian statistics (see next page).
2. **Time-based (duration in hours, cycles, miles)**: For example, item (e.g., tire) uniquely identified as x operated successfully for y hours under conditions z until it failed. In short, item x failed in y hours (z is not needed if the same for all items in the data set). A data set consisting of items with this type of **continuous** life data is often modeled with the *Weibull* probability distribution. A special case of the Weibull is when the failure rate is constant over time. Constant failure rate (λ) can be calculated using:
   a. $\lambda = \frac{failure\ count}{total\ run\ time}$ based on classical statistics.
   b. $\lambda = \frac{failure\ count+0.5}{total\ run\ time}$ based on one version of Bayesian statistics (see next page).
3. **Failure due to variation**: In this case, item x failed not as a function of time but due to **static stress**. That is, the item failed because its stress (load) exceeded its strength (capacity). The *Stress-Strength Interference* method calculates the probability of failure being the area described by the intersection of the stress distribution and the strength distribution. Note: A **safety factor** or safety margin are not sufficient to address failures due to the variation in the item's stress and the strength!

 Tim.Adams@NASA.gov

# Risk and Failure Probability, What is it?

## Failure Rate Formulas Based on Bayesian Statistics[1]

| Data Type | Demand Based<br>(failure on demand) | Time Based<br>(failure while operating) |
|---|---|---|
| **Failure Rate Formula[2,3]** | $p = \frac{failure\ count + 0.5}{total\ number\ of\ attempts + 1}$ | $\lambda = \frac{failure\ count + 0.5}{total\ run\ time}$ |
| **Prior Distribution[4]** | Beta distribution with $\alpha_{prior}$ = 0.5 and $\beta_{prior}$ = 0.5 being a Jeffreys Prior | Gamma distribution with $\alpha_{prior}$ = 0.5 and $\beta_{prior}$ = 0 being a Jeffreys Prior |
| **Likelihood Function** | Binomial distribution | Poisson distribution |
| **Posterior Distribution[5]** | Beta distribution with parameters $\alpha_{post} = x + \alpha_{prior}$ and $\beta_{post} = n - x + \beta_{prior}$ where x is failure count and n is number of demands. The mean of the beta distribution is $\frac{\alpha}{\alpha+\beta}$. | Gamma distribution with parameters $\alpha_{post} = x + \alpha_{prior}$ and $\beta_{post} = t + \beta_{prior}$ where x is failure count and t is total run time. The mean of the gamma distribution is $\frac{\alpha}{\beta}$. |
| **NASA PRA Guidebook[6]** | Page C-6 (pdf page 363) | Page C-11 (pdf page 369) |
| **NASA Handbook on Bayesian Inference[7]** | Page 34 (pdf page 54) | Page 40 (pdf page 60) |

**Endnotes:**

[1] **Bayesian statistics** quantitatively combines human belief (a subjectively-based probability distribution) with operational or test data (an objectively-based probability distribution).

[2] When the *failure count is zero*, these two Bayesian-based formulas are commonly used.

[3] When the *failure count is zero* and the data type is time-based, one method in classical statistics calculates the failure rate using: $\lambda = \frac{1/3}{total\ run\ time}$.

[4] A **Jeffreys Prior** is used when there is insufficient information to form an informed prior distribution. Thus, the Jefferys Prior is referred to as a noninformative prior and is intended to convey little prior belief or information. A **noninformative prior** allows the data (described by the likelihood function) to speak for themselves.

[5] A Bayesian-based failure-rate formula is the mean (average) of its posterior distribution. This mean is commonly called the point Bayes' estimate. A **posterior distribution** is derived from Bayes' Theorem (Bayes-Laplace Theorem). This Theorem uses a **prior distribution** (to represent the value of the failure rate as a belief or best estimate prior to collecting field data) and a **likelihood function** (the failure distribution for field data that was collected after the stated belief). The posterior distribution is shifted in the direction of the likelihood function that was used.

[6] Source: http://www.hq.nasa.gov/office/codeq/doctree/SP20113421.pdf

[7] Source: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090023159.pdf

 Tim.Adams@NASA.gov